修士論文 要旨

# A Study on Music Similarity by Recurrence Plots

# リカレンスプロットを用いた楽曲の類似度に関する研究

数理情報学専攻 66212　**澤井 賢一**

指導教員　**合原 一幸 教授**

## 1　はじめに

近年の科学技術の発達により，個人が所有，アクセスできる情報の量は急速に増えている．またそれに伴い，文字情報だけでなく，音声や映像情報も多く流通するようになった．そのため音楽情報処理の研究は，現在注目される研究分野の一つとなっている．その中でも類似度に関しては，音楽情報検索システムとの関わりから，多くの研究が行われている．

しかし，従来の類似度は短いフレーズが対象であり，楽曲全体の類似度を測ることには不向きであった．本研究では，楽譜に時系列データ解析のための手法を適用することで，楽曲全体の情報を反映させて類似度を測る手法を提案する．

## 2　リカレンスプロット

リカレンスプロットは，本来複雑な時系列データを視覚的に解析するために提案された手法である [1]．特に本研究では，リカレンスプロットから派生したクロスリカレンスプロットという手法を扱う．二つの時系列 $\boldsymbol{x}_i, \boldsymbol{y}_j$ $(i = 1, 2, \ldots, M,\ j = 1, 2, \ldots, N)$ に対するクロスリカレンスプロットとは，時刻 $i, j$ におけるそれぞれのデータの値がある閾値より近いときに，平面上に点 $(i, j)$ をプロットしたものである．具体的には次のように定義される：

$$R_{ij}(\epsilon) = \begin{cases} 1 & (||\boldsymbol{x}_i - \boldsymbol{y}_j|| < \epsilon), \\ 0 & \text{otherwise.} \end{cases}$$

ただし $\epsilon$ は閾値で，また，$R_{ij}(\epsilon) = 1$ なる $(i, j)$ がプロットする点を表す．

ある二つの時系列データに対するクロスリカレンスプロットには，二つのデータが近い値を取りやすいとき，取りにくいときに比べてより多くの点がプロットされる．そのため，プロットされうる点の数に対する実際にプロットされた点の数の割合は，二つの時系列データの類似度を測る指標とされている．

また，二つのデータ値が同じような推移をすると，たとえば $(i, j), (i+1, j+1), \ldots, (i+l, j+l)$ がプロットされ，対角方向と平行な線分が現れる．このため，プロットされた点の数に対する対角線と平行な線分をなす点の数の割合は，二つの時系列を生成する法則の類似度を示す指標として用いられている．

## 3　リカレンスプロットによる楽曲間類似度

本研究では，$i$ 番目から $i+n-1$ 番目までの $n$ 個の音符の列を，時刻 $i$ におけるデータとみなす．そして，$n$ 個の音符列間に距離を定めることで，リカレンスプロットを楽譜に適用する．本研究では $n$ 個の音符列に対する距離関数として二つのものを提案し，それぞれ本節と次節で論じる．

一つ目の手法では，音高とリズムの情報をもとに音符列間の類似度を定義し，これを距離関数と見る．この方法で，三つの楽曲に対してクロスリカレンスプロットを適用し，前節で述べた特徴量の値を比較した．その結果，より似ていないとした楽曲の組に比べ，より似ていると感じる楽曲の組のクロスリカレンスプロットの二つの特徴量は，いずれも大きな値となった．

## 4　$n$-グラムモデルとリカレンスプロットの関係

本節では前節とは異なり，音高の情報のみを用いる．また，同じ音名の音は同一視して扱う．すなわち，高音域にある「ド」も低音域にある「ド」も同じものとみなし，扱う音高は $\{$ ド, ド$\sharp$, $\ldots$, シ $\}$ の 12 種類とする．そして，二つの音符列間の距離は，音高の列が一致したら 0，一カ所でも音高が異なれば $\infty$ とする．このような距離のもとで二つの楽曲間のクロスリカレンスプロットを考えると，プロットされる点の数は，長さ $n$ のパターンの一致具合を表す．このクロスリカレンスプロットは，$n$-グラムモデルを用いた判別問題と関連付けることができる．

$n$-グラムモデルは，もともと自然言語処理の分野で広く用いられてきた手法である．$n$-グラムとは，文章中に含まれる長さ $n$ の部分文字列のことで，各 $n$-グラムの頻度を数え，その分布から文章の特徴を抽出するのが $n$-グラムモデルである．この $n$-グラムモデルとリカレンスプロットの関係を説明するにあたり，まず一般的な線形判別問題について述べる．ここでは，ある楽曲に

おける $n$-グラムの頻度ベクトルを楽曲の特徴ベクトルと考える．そして，特徴ベクトルに対する線形判別関数を訓練用の楽譜から算出する．

ベクトル $\boldsymbol{x}$ に対する線形判別関数 $f(\boldsymbol{x}) = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}$ は，値の正負によって $\boldsymbol{x}$ が二つのクラスのどちらに属するかを判別する．ここでは関数を特徴づける定ベクトル $\boldsymbol{\beta}$ を，訓練用の楽譜から最小二乗法で求めることを考える．すなわち，訓練用楽譜を $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$，それらを横に並べた行列を $X$，第 $i$ 成分が $\boldsymbol{x}_i$ の属するクラスを表した教師信号ベクトルを $\boldsymbol{y} \in \{-1, 1\}^N$ として，

$$\min_{\boldsymbol{\beta}}. \ ||\boldsymbol{y} - X^{\mathrm{T}}\boldsymbol{\beta}||^2 + \lambda||\boldsymbol{\beta}||^2,$$

を解く．ただし $\lambda$ は，判別関数が訓練用楽譜に対して過学習を起こさないためのパラメータである．この問題の解は解析的に，

$$\boldsymbol{\beta} = (XX^{\mathrm{T}} + \lambda \mathrm{I})^{-1}X\boldsymbol{y},$$

と求まる．この問題は，変数変換 $\boldsymbol{\beta} = X\boldsymbol{\alpha}$ を行っても等価であることが知られている．この変換により判別関数は $f(\boldsymbol{x}) = \boldsymbol{\alpha}^{\mathrm{T}}X^{\mathrm{T}}\boldsymbol{x}$ となる．また解くべき問題は，訓練用楽譜の特徴ベクトルのグラム行列を $K = X^{\mathrm{T}}X$ と書くと

$$\min_{\boldsymbol{\alpha}}. \ ||\boldsymbol{y} - K\boldsymbol{\alpha}||^2 + \lambda\boldsymbol{\alpha}^{\mathrm{T}}K\boldsymbol{\alpha},$$

となり，その解は

$$\boldsymbol{\alpha} = (K + \lambda \mathrm{I})^{-1}\boldsymbol{y},$$

となる．ここで，判別関数に現れるベクトル $(X^{\mathrm{T}}\boldsymbol{x})$ は，判別したい楽譜の特徴ベクトル $\boldsymbol{x}$ と楽譜 $\boldsymbol{x}_i$ との内積を第 $i$ 成分に持つベクトルで，行列 $K$ も訓練用楽譜の特徴ベクトル同士の内積が成分である．すなわち，この変数変換を行うことで，各特徴ベクトルが陽に分かっていなくとも，お互いの内積が別に計算できれば，この判別問題を考えることができる．今考えている特徴ベクトルは大きさが $12^n$ で，行列 $X$ を含む演算には大きな空間計算量が必要となる．しかし，上のような変数変換により，計算すべき問題の規模を小さくすることができる．このように，高次元の特徴空間を介さずに内積の値を用いることで問題の規模を縮小化することは，一般にカーネルトリックと呼ばれている．

次に，本節で考えるリカレンスプロットにより，特徴ベクトル間の内積が求まることを示す．二つの楽曲 $s_i, s_j$ に対する特徴ベクトルを $\boldsymbol{x}_i, \boldsymbol{x}_j$ とする．また，長さ $n$ の音高パターンの集合を $P = \{p_1, p_2, \ldots, p_q\}$ $(q = 12^n)$ とし，楽曲 $s_i, s_j$ 中にパターン $p$ が出現する回数をそれぞれ $n_i(p), n_j(p)$ で表す．このとき二つのベクトルの内積は，$\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j = \sum_{p \in P} n_i(p)n_j(p)$ と表わされる．この二つの楽曲に対するクロスリカレンスプロットにおいて，パターン $p$ に一致したことによる点だけ取り出すと，二辺が $n_i(p), n_j(p)$ 個の点からなる長方形の格子となる．そして，その長方形をなす点の数は $n_i(p)n_j(p)$ である．すべてのパターンに対してそれぞれに対する長方形格子を考えれば，このクロスリカレンスプロットに現れる点の総数は $\sum_{p \in P} n_i(p)n_j(p)$ と表わされる．すなわち，特徴ベクトルの内積と一致する．

特徴ベクトルとして成分の和が 1 となるよう正規化したものを考えれば，特徴ベクトルの内積はクロスリカレンスプロットでプロットされる点の数の割合に対応する．この判別問題を 3 人の作曲家に対して適用したときの，部分列の長さ $n$ と正答率の関係を図 1 に示す．判別自体はリカレンスプロットを用いなくてもできるが，リカレンスプロットを用いることで計算量が減り，より大きな $n$ に対する問題も容易に扱うことができる．正答率の算出は，クロスバリデーションにより行った．この結果より，作曲家 $\mathrm{C}_1$ は $n$-グラムモデルにより特徴が表せるという結論が得られた．

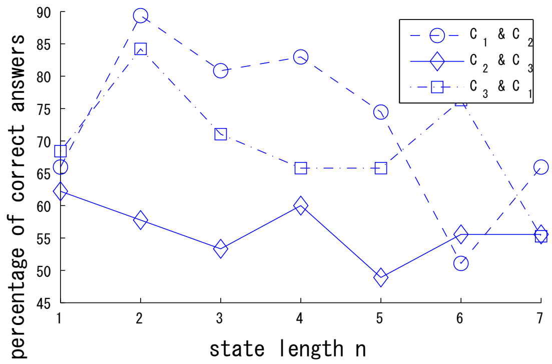


**図** 1. 3 人の作曲家の判別の結果.

## 5 おわりに

1-グラムに対応するリカレンスプロットにおいて，対角方向と平行な長さ 3 の線分は， 3-グラムの一致があったことを意味する．つまり， 1-グラムのリカレンスプロットには $\forall n \geq 2$ に対する $n$-グラムの情報が含まれる．すべての $n$ に対する情報を包括的に扱う手法を提案することが，今後の課題の一つである．

## 参考文献

[1] Jean-Pierre Eckmann, Sylvie Oliffson Kamphorst, and David Ruelle. Recurrence plots of dynamical systems. *Europhysics Letters*, 4(9):973–977, 1987.